

Guidance for Industry

Evidence-Based Review System for the Scientific Evaluation of Health Claims

Additional copies are available from:
Office of Nutrition, Labeling and Dietary Supplements
Division of Nutrition Programs and Labeling HFS-830
Center for Food Safety and Applied Nutrition
Food and Drug Administration
5100 Paint Branch Parkway
College Park, MD 20740
(Tel) 301-436-1450
<http://www.cfsan.fda.gov/guidance.html>

You may submit written or electronic comments regarding this guidance at any time. Submit written comments on the guidance to the Division of Dockets Management (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. Submit electronic comments to <http://www.regulations.gov>. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Food Safety and Applied Nutrition
January 2009

FDA-2007-D-0371

GDL

Table of Contents

I. Introduction

II. Background

III. Evidence-Based Review System for the Scientific Evaluation of Health Claims

IV. References

Guidance for Industry¹

Evidence-Based Review System for the Scientific Evaluation of Health Claims

This guidance represents the Food and Drug Administration's (FDA's) current thinking on this topic. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. You can use an alternative approach if the approach satisfies the requirements of the applicable statutes and regulations. If you want to discuss an alternative approach, contact the FDA staff responsible for implementing this guidance. If you cannot identify the appropriate FDA staff, call the appropriate telephone number listed on the title page of this guidance.

I. Introduction

This guidance document is for industry. It represents the agency's current thinking on 1) the process for evaluating the scientific evidence for a health claim, 2) the meaning of the significant scientific agreement (SSA) standard in section 403(r)(3) of the Federal Food, Drug, and Cosmetic Act (the Act) (21 U.S.C. 343(r)(3)) and 21 CFR 101.14(c), and 3) credible scientific evidence to support a qualified health claim.

This guidance document describes the evidence-based review system that FDA intends to use to evaluate the publicly available scientific evidence for SSA health claims or qualified health claims on the relationship between a substance and a disease or health-related condition.² This guidance document explains the agency's current thinking on the scientific review approach FDA should use and is intended to provide guidance to health claim petitioners.³

¹ This guidance has been prepared by the Office of Nutrition, Labeling, and Dietary Supplements in the Center for Food Safety and Applied Nutrition at the U.S. Food and Drug Administration.

² For brevity, "disease" will be used as shorthand for "disease or health-related condition". "Disease or health-related condition" is defined as damage to an organ, part, structure, or system of the body such that it does not function properly (e.g., cardiovascular disease), or a state of health leading to such dysfunctioning (e.g., hypertension). 21 CFR 101.14(a)(5).

³ This new guidance document replaces FDA's guidance entitled "Guidance for Industry and FDA: Interim Evidence-based Ranking System for Scientific Data," which addressed the scientific review of qualified health claims. Although the interim evidence-based ranking system guidance included a section on ranking

Contains Nonbinding Recommendations

The specific topics addressed in this guidance document are: (1) identifying studies that evaluate the substance/disease relationship, (2) identifying surrogate endpoints for disease risk, (3) evaluating the human studies to determine whether scientific conclusions can be drawn from them about the substance/disease relationship, (4) assessing the methodological quality of each human study from which scientific conclusions about the substance/disease relationship can be drawn, (5) evaluating the totality of scientific evidence, (6) assessing significant scientific agreement, (7) specificity of claim language for qualified health claims, and (8) reevaluation of existing SSA or qualified health claims.

FDA's guidance documents, including this guidance, do not establish legally enforceable responsibilities. Instead, guidances describe the Agency's current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word *should* in Agency's guidances means that something is suggested or recommended, but not required.

II. Background

The Nutrition Labeling and Education Act of 1990 (NLEA) (Pub. L. 101-553) was designed to give consumers more scientifically valid information about foods they eat. Among other provisions, the NLEA directed FDA to issue regulations providing for the use of statements that describe the relationship between a substance and a disease ("health claims") in the labeling of foods, including dietary supplements, after such statements have been reviewed and authorized by FDA.⁴ For these health claims, that is, statements about substance/disease relationships, FDA has defined the term "substance" by regulation as a specific food or food component (21 CFR 101.14(a)(2)). An authorized health claim may be used on both conventional foods and dietary supplements, provided that the substance in the product and the product itself meet the appropriate standards in the authorizing regulation. Health claims are directed to the general population or designated subgroups (e.g., the elderly) and are intended to assist the consumer in maintaining healthful dietary practices.

In evaluating a petition for an authorized health claim, FDA considers whether the evidence supporting the relationship that is the subject of the claim meets the SSA

the strength of the scientific evidence, this new guidance document does not include such a section because studies are being conducted on the consumer's understanding of various possible ranking systems that could be used to describe the strength of the evidence for a health claim. FDA intends to reexamine its ranking systems and issue appropriate guidance once these studies are completed. In addition, this guidance document replaces FDA's guidance entitled "Guidance for Industry: Significant Scientific Agreement in the Review of Health Claims for Conventional Foods and Dietary Supplements."

⁴ In 1997, Congress enacted the Food and Drug Administration Modernization Act, which established an alternative authorization procedure for health claims based on authoritative statements of certain federal scientific bodies or the National Academy of Sciences. This guidance document does not address that alternative procedure.

Contains Nonbinding Recommendations

standard. This standard derives from 21 U.S.C. 343 (r)(3)(B)(i), which provides that FDA shall authorize a health claim to be used on conventional foods if the agency “determines based on the totality of the publicly available evidence (including evidence from well-designed studies conducted in a manner which is consistent with generally recognized scientific procedures and principles), that there is significant scientific agreement among experts qualified by scientific training and experience to evaluate such claims, that the claim is supported by such evidence.” This scientific standard was prescribed by statute for conventional food health claims; by regulation, FDA adopted the same standard for dietary supplement health claims. See 21 CFR 101.14(c).

The genesis of qualified health claims was the court of appeals decision in *Pearson v. Shalala* (*Pearson*). In that case, the plaintiffs challenged FDA’s decision not to authorize health claims for four specific substance-disease relationships in the labeling of dietary supplements. Although the district court ruled for FDA (14 F. Supp. 2d 10 (D.D.C. 1998), the U.S. Court of Appeals for the D.C. Circuit reversed the lower court’s decision (164 F.3d 650 (D.C. Cir.1999)). The appeals court held that the First Amendment does not permit FDA to reject health claims that the agency determines to be potentially misleading unless the agency also reasonably determines that a disclaimer would not eliminate the potential deception. The appeals court also held that the Administrative Procedure Act (APA) required FDA to clarify the “significant scientific agreement” (SSA) standard for authorizing health claims.

On December 22, 1999, FDA announced the issuance of its *Guidance for Industry: Significant Scientific Agreement in the Review of Health Claims for Conventional Foods and Dietary Supplements* (64 Fed. Reg. 17494). This guidance document was issued to clarify FDA’s interpretation of the SSA standard in response to the court of appeals’ second holding in *Pearson*.

On December 20, 2002, the agency announced its intention to extend its approach to implementing the *Pearson* decision to include health claims for conventional foods (67 Fed. Reg. 78002). Recognizing the need for a scientific framework for qualified health claims, the Task Force on “Consumer Health Information for Better Nutrition” was formed. The Task Force recognized that there could be significant public health benefits when consumers have access to, and use, more and better information in conventional food as well as dietary supplement labeling to aid them in their purchases, information that goes beyond just price, convenience, and taste, but extends to include science-based health factors. Armed with more scientifically based information about the likely health benefits of the foods and dietary supplements they purchase, consumers can make a tangible difference in their own long-term health by lowering their risk of numerous chronic diseases.

Contains Nonbinding Recommendations

To maximize the public health benefit of FDA's claims review process, the Task Force's Final Report⁵ provides a procedure to prioritize on a case-by-case basis all complete petitions according to several factors, including whether the food or dietary supplement that is the subject of the petition is likely to have a significant impact on a serious or life-threatening illness; the strength of the evidence; whether consumer research has been provided to show the claim is not misleading; whether the substance that is the subject of the claim has undergone an FDA safety review (i.e., is an authorized food additive, has been Generally Recognized as Safe (GRAS) affirmed, listed, or has received a letter of "no objection" to a GRAS notification); whether the substance that is the subject of the claim has been adequately characterized so that the relevance of available studies can be evaluated; whether the disease is defined and evaluated in accordance with generally accepted criteria established by a recognized body of qualified experts; and whether there has been prior review of the evidence or the claim by a recognized body of qualified experts.

As part of the Task Force's final report, FDA developed an interim evidence-based review system that the agency intended to use to evaluate the substance/disease relationships that are subjects of qualified health claims. In reviewing the December 22, 1999 SSA guidance document and the 2003 Task Force report, it became apparent to the agency that the components of the scientific review process for an SSA health claim and qualified health claim are very similar. Because of the similarity between the scientific reviews for SSA and qualified health claims, FDA intends to use the approach set out in this guidance for evaluating the scientific evidence in petitions that are submitted for an SSA health claim or qualified health claim. The evidence-based review system set out in this guidance will assist the agency in determining whether the scientific evidence meets the SSA standard or, if not, whether the evidence supports a qualified health claim. In addition to a science review, health claims undergo a regulatory review. Health claims that meet the SSA standard are authorized by publication of a final rule or an interim final rule in the Federal Register. For qualified health claims supported by credible evidence, FDA issues a letter regarding its intent to consider enforcement discretion.

Although this guidance replaces the *Guidance for Industry: Significant Scientific Agreement in the Review of Health Claims for Conventional Foods and Dietary Supplements* (64 Fed. Reg. 17494), issued to clarify FDA's interpretation of the SSA standard in response to the court of appeals' second holding in *Pearson*, FDA believes this guidance continues to be consistent with the court's holding. The basic principles of SSA articulated in the 1999 guidance have not changed. A finding of SSA still requires the agency's best judgment as to whether qualified experts would likely agree that the scientific evidence supports the substance/disease relationship that is the subject of a proposed health claim. In fact, many of the explanations of SSA in this guidance are taken verbatim from the 1999 guidance. This guidance represents further scientific

⁵ See guidance (Attachment A) entitled "Guidance for Industry and FDA: Interim Procedures for Qualified Health Claims in the Labeling of Conventional Human Food and Human Dietary Supplements," July 10, 2003 (<http://www.cfsan.fda.gov/~dms/hclmgui3.html>)

Contains Nonbinding Recommendations

developments in the agency's approach to the review of scientific evidence rather than a change in its understanding of what constitutes SSA.

III. Evidence-Based Review System for the Scientific Evaluation of Health Claims

A. What is an Evidence-Based Review System?

An evidence-based review system is a systematic science-based evaluation of the strength of the evidence to support a statement. In the case of health claims, it evaluates the strength of the scientific evidence to support a proposed claim about a substance/disease relationship. The evaluation process involves a series of steps to assess scientific studies and other data, eliminate those from which no conclusions about the substance/disease relationship can be drawn, rate the remaining studies for methodological quality and evaluate the strength of the totality of scientific evidence by considering study types, methodological quality, quantity of evidence for and against the claim (taking into account the numbers of various types of studies and study sample sizes), relevance to the U.S. population or target subgroup, replication of study results supporting the proposed claim, and overall consistency of the evidence. After assessing the totality of the scientific evidence, FDA determines whether there is SSA to support an authorized health claim, or credible evidence to support a qualified health claim.

B. Identifying Studies That Evaluate the Substance/Disease Relationship

The agency considers the publicly available data and written information pertaining to the relationship between a substance and disease. FDA reviews studies that must be submitted in petitions seeking health claims (21 CFR 101.70). Through a literature search, the agency identifies additional studies that are relevant to the proposed health claim. Before the strength of the evidence for a substance/disease relationship can be assessed, FDA separates individual relevant articles on human studies from other types of data and information. FDA intends to focus its review primarily on articles reporting human intervention and observational studies because only such studies can provide evidence from which scientific conclusions can be drawn about the substance/disease relationship in humans. Next, the agency considers a number of threshold questions in the review of the scientific evidence:

- *Have the studies specified and measured the substance that is the subject of the claim?* Studies should identify a substance that is measurable. A "substance" is defined as a specific food or component of food regardless of whether the food is in conventional food form or a dietary supplement. 21 CFR 101.14(a) (2). A food component can be, for example, a nutrient or dietary ingredient.⁶ If the substance is to be consumed as a component of conventional food at decreased dietary levels, the substance must be a

⁶ See 21 U.S.C. 321(ff)(1).

Contains Nonbinding Recommendations

nutrient that is required to be included in the Nutrition Facts label (21 CFR 101.14(b)(2)). If the substance is to be consumed at other than decreased dietary levels, the substance must contribute taste, aroma, nutritive value,⁷ or a technical effect listed in 21 CFR 170.3(o) to the food, and must be safe and lawful for use at the levels necessary to justify a claim (21 CFR 101.14(b)(3)).

- *Have the studies appropriately specified and measured the specific disease or health-related condition that is the subject of the claim?* "Disease or health-related condition" is defined as damage to an organ, part, structure, or system of the body such that it does not function properly (e.g., cardiovascular disease), or a state of health leading to such dysfunctioning (e.g., hypertension). 21 CFR 101.14(a) (5). Studies should identify a specific measurable disease or health-related condition by either measuring incidence, associated mortality, or validated surrogate endpoints that predict risk of a specific disease.

For example, cancer is a constellation of more than 100 different diseases, each characterized by the uncontrolled growth and spread of abnormal cells (American Cancer Society, 2004). Cancer is categorized into different types of diseases based on the organ and tissue sites (National Cancer Institute). Cancers at different organ sites have different risk factors, treatment modalities, and mortality risk (American Cancer Society, 2004). Both genetic and environmental (including diet) risk factors may affect the risk of different types of cancers. Risk factors may include a family history of a specific type of cancer, cigarette smoking, alcohol consumption, overweight and obesity, exposure to ultraviolet or ionizing radiation, exposure to cancer-causing chemicals, and dietary factors. The etiology, risk factors, diagnosis, and treatment for each type of cancer are unique (Hord et al., 2007; Milner et al., 2006). Since each form of cancer is a unique disease based on organ site, risk factors, treatment options, and mortality risk, FDA's current approach is to evaluate each form of cancer individually in a health claim or qualified health claim petition to determine whether the scientific evidence supports the potential substance-disease relationship for that type of cancer, which would constitute a disease under 21 CFR 101.14(a)(5). The agency has used this approach in several letters of enforcement discretion including green tea and cancer dated June 30, 2005 (<http://www.cfsan.fda.gov/~dms/qhc-gtea.html>), tomatoes/lycopene and various cancers dated November 8, 2005 (<http://www.cfsan.fda.gov/~dms/qhclyco.html>), calcium and various cancers dated October 12, 2005 (<http://www.cfsan.fda.gov/~dms/qhcca2.html>) as well as the Federal Register notice entitled "Health Claims and Qualified Health Claims; Dietary Lipids and Cancer, Soy Protein and Coronary Heart Disease, Antioxidant Vitamins and Certain Cancers, and Selenium and Certain Cancers; Reevaluation" (72 Fed. Reg. 72738, December 21, 2007)

After considering these threshold issues, FDA categorizes the studies by type.

⁷ "Nutritive value" is defined in 21 CFR 101.14(a)(3) as value in sustaining human existence by such processes as promoting growth, replacing loss of essential nutrients, or providing energy.

Contains Nonbinding Recommendations

• Intervention Studies

In an intervention study, subjects are provided the substance (food or food component) of interest (intervention group), typically either in the form of a conventional food or dietary supplement. The quality and quantity of the substance should be controlled for. In randomized controlled trials, subjects are assigned to an intervention group by chance rather than systematically. Individual subjects may not be similar to each other, but the intervention and control groups should be similar after randomization. Randomized controlled trials offer the best assessment of a causal relationship between a substance and a disease because they control for known confounders of results (i.e., other factors that could affect risk of disease). Through random assignment of subjects to the intervention and control groups, these studies avoid selection bias -- that is, the possibility that those subjects most likely to have a favorable outcome, independent of an intervention, are preferentially selected to receive the intervention. Potential bias is also reduced by "blinding" the study so that the subjects do not know whether they are receiving the intervention, or "double blinding," in which neither the subjects nor the researcher who assesses the outcome knows who is in the intervention group and who is in the control group. By controlling the test environment, including the amount and composition of substance consumed and all other dietary factors, these studies also can minimize the effects of variables or confounders on the results.⁸ Therefore, randomized, controlled intervention studies provide the strongest evidence of whether or not there is a relationship between a substance and a disease (Greer et al., 2000).

Furthermore, such studies can provide convincing evidence of a cause and effect relationship between an intervention and an outcome (Kraemer et al., 2005 at 113). Randomization, however, may result in unequal distribution of the characteristics of the subjects between the control and treatment groups (e.g., baseline age or blood [serum or plasma] LDL cholesterol levels are significantly different). If the baseline values are significantly different, then it is difficult to determine if differences at the end of the study were due to the intervention or to differences at the beginning of the study. When the substance is provided as a supplement, a placebo should be provided to the control group. When the substance is a food, it may not be possible to provide a placebo and therefore subjects in such a study may not be blinded. Although the study may not be blinded in this case, a control group is still needed to draw conclusions from the study.

Randomized controlled trials typically have either a parallel or cross-over design. Parallel design studies involve two groups of subjects, the test group and the control group, which simultaneously receive the substance or serve as the control, respectively. Cross-over design involves all subjects crossing over from the intervention group to the control group, and vice versa, after a defined time period.

⁸ Confounders are factors that are associated with both the disease in question and the intervention, and that if not controlled for, prevent an investigator from being able to conclude that an outcome was caused by an intervention.

Contains Nonbinding Recommendations

Although intervention studies are the most reliable category of studies for determining a cause-and effect relationship, generalizing from the studies conducted on selected populations to different populations may not be scientifically valid. For example, if the evidence consists of studies showing an association between intake of a substance and reduced risk of juvenile diabetes, then such studies should not be extrapolated to the risk of diabetes in adults.

• Observational Studies

Observational studies measure associations between the substance and disease. Observational studies lack the controlled setting of intervention studies. Observational studies are most reflective of free-living⁹ populations and may be able to establish an association between the substance and the disease. In contrast to intervention studies, observational studies cannot determine whether an observed relationship represents a relationship in which the substance caused a reduction in disease risk or is a coincidence (Sempos et al., 1999). Because the subjects are not randomized based on various disease risk factors at the beginning of the study, known confounders of disease risk need to be collected and adjusted for to minimize bias. For example, information on each subject's risk factors, such as age, race, body weight and smoking, should be collected and used to adjust the data so that the substance/disease relationship is accurately measured. Risk factors that need to be adjusted for are determined for each disease being studied. For example, the risk of cardiovascular disease increases with age; therefore, an adjustment for age is needed in order to eliminate potential confounding.

In determining whether the substance that is the subject of the claim has been measured appropriately, it is important to critically evaluate the method of assessment of dietary intake. Many observational studies rely on self-reports of diet (e.g., diet records, 24-hour recalls, diet histories, and food frequency questionnaires), which are estimates of food intake (National Research Council, 1989). Diet records are based on the premise that food weights provide an accurate estimation of food intake. Subjects weigh the foods they consume and record those values. The 24-hour recall method requires that subjects describe which foods and how much of each food they consumed during the prior 24-hour period. Diet histories use questionnaires or interviewers to estimate the typical diet of subjects over a certain period of time. A food frequency questionnaire is the most common dietary assessment tool used in large observational studies of diet and health. Validated food frequency questionnaires are more reliable in estimating "usual" intake of foods than diet records or 24-hour recall methods (Subar et al., 2001). The questionnaire asks participants to report the frequency of consumption and portion size from a list of foods over a defined period of time. One problem with the dietary intake assessment methods described above is that there may be bias in the self-reporting of certain foods. For example, individuals who are overweight tend to under-report their portion sizes (Flegal et al., 1999) and therefore the actual amount of substances consumed is often

⁹ Free-living populations represent those who consume diets and have lifestyles (e.g., smoking, drinking, and exercise) of their own choice.

Contains Nonbinding Recommendations

underestimated. If there are reliable biomarkers of intake¹⁰ of a substance, these biomarkers are often measured rather than using self-reported intakes.

Observational studies may be prospective or retrospective. These types of studies are subject to different forms of bias (information and selection).¹¹ In prospective studies, investigators recruit subjects and observe them prior to the occurrence of the disease outcome. Prospective observational studies compare the incidence of a disease with exposure to the substance. In retrospective studies, investigators review the medical records of subjects and/or interview subjects after the disease has occurred. Retrospective studies are particularly vulnerable to measurement error and recall bias because they rely on subjects' recollections of what they consumed in the past. Because of the limited ability of observational studies to control for variables, they are often susceptible to confounders, such as complex substance/disease interactions.

Well-designed observational studies can provide useful information for identifying possible associations to be tested by intervention studies (Kraemer et al., 2005 at 107). In contrast to intervention studies, even the best-designed observational studies cannot establish cause and effect between an intervention and an outcome (Kraemer et al., 2005 at 114). However, as discussed above, intervention studies can test whether there is evidence to show a cause and effect between a substance and a reduced risk of a disease. Observational studies from which scientific conclusions can be drawn, in some situations, can be support for a substance/disease relationship for an SSA or qualified health claim. Each observational study design has its strength and weaknesses as discussed below (Sempos et al., 1999).

Cohort studies are prospective studies that compare the incidence of a disease in subjects who receive a specific exposure of the substance that is the subject of the claim with the incidence of the disease in subjects who do not receive that exposure. Because the intake of the substance precedes disease development, this study design ensures that the subjects are not consuming the substance in response to having the disease. Cohort studies can yield relative estimates of risk (Szklo and Nieto, 2000).¹² Cohort studies are considered to be the most reliable observational study design (Greer et al., 2000).

In *case-control studies*, subjects with a disease (cases) are compared to subjects who do not have the disease (controls).¹³ Prior intake of the substance is estimated from dietary assessment methods for both cases and control. These retrospective studies often ask about food consumption at least 1 year prior to diagnosis of the disease, making it

¹⁰ Biomarkers of intake are measurements of the substance itself or a metabolite of the substance in biological samples (e.g., serum selenium) that have been validated to confirm that they reflect the intake of that substance.

¹¹ Bias is the systematic error that may result in flaws from subject selection (selection bias) or exposure and disease outcome measurements (information bias) (Szklo and Nieto, 2000).

¹² Relative risk is expressed as the ratio of the risk (disease incidence) in exposed individuals to that in unexposed individuals. It is calculated in prospective cohorts by measuring exposure of the substance in subjects with and without disease. An adjusted relative risk controls for potential confounders.

¹³ An example of a case-control study is a study design that assesses parameters related to the frequency and distribution of disease in a population, such as leading cause of death.

Contains Nonbinding Recommendations

difficult to obtain an accurate estimate of intake. Furthermore, a key assumption is that food consumption has not been altered by the disease process or by knowledge of having the disease. Thus, the case-control study design does not control for changes in intake caused by or in response to the disease. Case-control studies can yield an odds ratio, which is an estimate of the relative risk of getting the disease (Szklo and Nieto, 2000).¹⁴ Case-control studies are considered to be less reliable than cohort studies (Greer et al., 2000).

A *nested-case control or case-cohort* study uses subjects from a pre-defined cohort, such as the population of an ongoing cohort study. Cases are subjects diagnosed with the disease (e.g., lung cancer) in the cohort. In a nested-case control study, controls are subjects selected from individuals at risk each time a case (e.g., lung cancer) is diagnosed. In a case-cohort study, controls are selected randomly from the baseline cohort (Szklo and Nieto, 2000). Either a relative risk or odds ratio may be calculated in these types of studies. Nested-case control or case-cohort studies are considered less reliable than cohort studies but more reliable than case-control studies.

Cross-sectional studies usually involve collecting information on food consumption at a single point in time in individuals with and without a specific disease.¹⁵ These studies can be useful for identifying possible correlates (i.e., by determining the correlation coefficient¹⁶ between dietary intake of a substance and prevalence of a disease) and for providing baseline information for subsequent prospective studies (Kraemer et al., 2005 at 99-100). However, because dietary intake and disease status are measured at the same time, it is not possible to determine whether dietary intake of the substance is a factor affecting the risk of the disease or a result of having the disease. Cross-sectional studies calculate the prevalence of a disease based on exposure and this may be a measure of survival of the disease rather than the risk of developing the disease (Szklo and Nieto, 2000). Further, cross-sectional studies are considered to be a "relatively weak method of studying diet-disease associations" because they can be subject to significant potential measurement error regarding dietary intake due to inaccuracy of survey methods used and limited ability to control for dietary intake variations (Sempos et al., 1999). For these reasons, cross-sectional study results "have the potential to mislead as errors of interpretation are very common" (Kraemer et al., 1005 at 103). Cross-sectional studies are considered to be less reliable than cohort and case-control studies (Greer et al., 2000).

¹⁴ An odds ratio is the odds of developing the disease in exposed compared to unexposed individuals. It is calculated in case control studies by measuring disease development in subjects based on exposure to the substance. Adjusted odds ratio controls for potential confounders.

¹⁵ A few cross-sectional studies are time-series studies that compare outcomes during different time periods (e.g., whether the rate of occurrence of a particular outcome during one five-year period changed during a subsequent five-year period).

¹⁶ The correlation coefficient (r) is a measure of the interdependence of two variables, such as intake of a substance and prevalence of a disease. It is expressed as a point on a scale of -1 and 1, where -1 indicates a perfect negative correlation, 0 indicates an absence of correlation, and +1 indicates a perfect positive correlation at +1. See Webster's II New Riverside University Dictionary (Riverside Publishing Co., 1984).] Thus, the closer the correlation coefficient is to either of the endpoints on the scale, the stronger the relationship between the two variables.

Contains Nonbinding Recommendations

Ecological studies compare disease incidence across different populations. *Case reports* describe observations of a single subject or a small number of subjects. Ecological studies and case reports are the least reliable types of observational studies.

• Research Synthesis Studies

Reports that discuss a number of different studies, such as review articles,¹⁷ do not provide sufficient information on the individual studies reviewed for FDA to determine critical elements such as the study population characteristics and the composition of the products used. Similarly, the lack of detailed information on studies summarized in review articles prevents FDA from determining whether the studies are flawed in critical elements such as design, conduct of studies, and data analysis. FDA must be able to review the critical elements of a study to determine whether any scientific conclusions can be drawn from it. Therefore, FDA intends to use review articles and similar publications¹⁸ to identify reports of additional studies that may be useful to the health claim review and as background about the substance/disease relationship. If additional studies are identified, the agency intends to evaluate them individually. Most meta-analyses,¹⁹ because they lack detailed information on the studies summarized, will only be used to identify reports of additional studies that may be useful to the health claim review and as background about the substance-disease relationship. FDA, however, intends to consider as part of its health claim review process a meta-analysis that reviews all the publicly available studies on the substance/disease relationship. The reviewed studies should be consistent with the critical elements, quality and other factors set out in this guidance and the statistical analyses adequately conducted.

• Animal and *in vitro* Studies

FDA intends to use animal and *in vitro* studies as background information regarding mechanisms that might be involved in any relationship between the substance and disease. The physiology of animals is different than that of humans. *In vitro* studies are conducted in an artificial environment and cannot account for a multitude of normal physiological processes such as digestion, absorption, distribution, and metabolism that affect how humans respond to the consumption of foods and dietary substances (IOM, 2005). Animal and *in vitro* studies can be used to generate hypotheses, investigate biological plausibility of hypotheses, or to explore a mechanism of action of a specific food component through controlled animal diets; however, these studies do not provide information from which scientific conclusions can be drawn regarding a relationship between the substance and disease in humans.

C. Identifying Surrogate Endpoints of Disease Risk

¹⁷ Review articles summarize the findings of individual studies on a given topic.

¹⁸ Other examples include book chapters, abstracts, letters, and committee reports.

¹⁹ A meta-analysis is the process of systematically combining and evaluating the results of clinical trials that have been completed or terminated (Spilker, 1991).

Contains Nonbinding Recommendations

Surrogate endpoints are risk biomarkers²⁰ that have been shown to be valid predictors of disease risk and therefore may be used in place of clinical measurements of the onset of the disease in a clinical trial (Spilker, 1991). Because a number of diseases develop over a long period of time, it may not be possible to carry out the study for a long enough period to see a statistically meaningful difference in the incidence of disease among study subjects in the treatment and control groups.

These are examples of surrogate endpoints of disease risk accepted by the National Institutes of Health and/or FDA's Center for Drug Evaluation and Research: (1) serum low-density lipoprotein (LDL) cholesterol concentration, total serum cholesterol concentration, and blood pressure for cardiovascular disease; (2) bone mineral density for osteoporosis; (3) adenomatous colon polyps for colon cancer; and (4) elevated blood sugar concentrations and insulin resistance for type 2 diabetes.

There can be multiple pathways to a specific disease, such as cardiovascular disease. Therefore, the accepted surrogate endpoints that are involved in a single pathway may not be applicable to certain substances that are involved in a different pathway. For example, the long chain omega-3 fatty acids generally have no effect on serum LDL cholesterol levels, and studies suggest that these fatty acids alter cardiovascular risk through a different pathway. Therefore, LDL cholesterol levels cannot be used in evaluating the relationship between the long chain omega-3 fatty acids and risk of cardiovascular disease.

D. Evaluating Human Studies

Under the evidence-based review approach set out in this guidance, FDA intends to evaluate each individual human study to determine whether any scientific conclusions about the substance/disease relationship can be drawn from the study. Certain critical elements of a study, such as design, data collection, and data analysis, may be so seriously flawed that they make it impossible to draw scientific conclusions from the study. FDA does not intend to use studies from which it cannot draw any scientific conclusions about the substance/disease relationship, and plans to eliminate such studies from further review. Below are examples of questions that the agency intends to consider whether scientific conclusions can be drawn from an intervention or observational study about the substance/disease relationship.

Intervention Studies

- *Were the study subjects healthy or did they have the disease that is the subject of the health claim?* Health claims involve reducing the risk of a disease in people who do not have the disease that is the subject of the claim. FDA considers evidence from studies with subjects who have the disease that is the subject of the claim only if it is scientifically appropriate to extrapolate to individuals who do not have the disease. That

²⁰ Risk biomarkers are biological indicators that signal a changed physiological state that is associated with the risk of a disease.

Contains Nonbinding Recommendations

is, the available scientific evidence demonstrates that (1) the mechanism(s) for the mitigation or treatment effects measured in the diseased populations are the same as the mechanism(s) for risk reduction effects in non-diseased populations and (2) the substance affects these mechanisms in the same way in both diseased and healthy people. If such evidence is not available, the agency cannot draw any scientific conclusions from studies that used subjects that have the disease that is the subject of the health claim to evaluate the substance/disease relationship and, therefore, the agency does not intend to use these studies to evaluate the substance/disease relationship. On the other hand, if, for example, FDA was reviewing a health claim on reduction of risk of coronary heart disease, it would consider studies that include individuals who have an unrelated disease (e.g., osteoporosis) or are at risk (e.g., elevated LDL cholesterol levels) of getting the disease that is the subject of the claim.

- *Was the disease that is subject of the claim measured as a “primary” endpoint?*

Intervention studies screen for prevalent cases of the disease at the beginning of the study to minimize bias. For example, intervention studies evaluating the recurrence of colorectal polyps prescreen the subjects to ensure there are no existing colorectal polyps at the onset of the intervention study. Intervention studies may evaluate the outcomes of other diseases as secondary endpoints, but do not screen for these diseases at the onset of the study. For example, a study evaluating the recurrence of colorectal polyps may also evaluate the incidence of prostate cancer; however, because the prostate cancer endpoint is not the primary endpoint, the study would not screen the subjects to ensure that they are free of prostate cancer before enrolling them. Consequently, the results with respect to prostate cancer may be biased due to an uneven distribution of cases of prostate cancer between the treatment and placebo groups at the beginning of the study. Uneven distribution of important patient or disease characteristics between groups may lead to mistaken interpretation (Spilker, 1991); therefore, scientific conclusions about a disease endpoint cannot be drawn from a study unless the study evaluates that outcome as a primary endpoint.

- *Did the study include an appropriate control group?* An appropriate control group represents study subjects who did not receive the substance. If an appropriate control group is not included, then it is not possible to ascertain whether changes in the endpoint of interest were due to the substance or due to unrelated and uncontrolled extraneous factors (Spilker, 1991; Federal Judicial Center, 2000). Without an appropriate control group, scientific conclusions cannot be drawn about a substance/disease relationship and, therefore, the agency does not intend to use these studies to evaluate the substance/disease relationship.

When the intervention study involves providing a whole food rather than a food component, the experimental and control diets should be similar enough that the relationship between the substance and disease can be evaluated. For example, if the substance is a specific type of fatty acid, then the composition of the experimental and control diets should be similar for all food components, except that particular fatty acid. Scientific conclusions cannot be drawn about the relationship between a substance and a disease when the amounts of other substances that are known to affect the risk of the

Contains Nonbinding Recommendations

disease that is subject of the claim are different between the control and experimental diets.

- *Was the study designed to measure the independent role of the substance in reducing the risk of a disease?* When the substance is a food component, it may not be possible to accurately determine its independent effects when whole foods or multi-nutrient supplements are provided to the intervention group. For example, if the claim is about a relationship between lutein and age-related macular degeneration (AMD), then scientific conclusions cannot be drawn from a study in which the intervention group received spinach or multi-nutrient supplements that contain other substances (e.g., vitamin C, vitamin E, and zinc) that have been suggested to have a role in protecting against AMD. As another example, if the substance is a fatty acid that has been shown to alter blood cholesterol levels, but the levels of other food components (e.g., cholesterol) known to affect cholesterol levels markedly vary between the intervention and control diets, then it is not possible to determine the independent effect of the fatty acid.

- *Were the relevant baseline data (e.g., on the surrogate endpoint) significantly different between the control and intervention group?* If the baseline values for the endpoint being measured are significantly different, then it is difficult to interpret the findings of the intervention. For example, in a study of the effects of a low-sodium diet on the risk of cardiovascular disease, having baseline blood pressure levels higher in the intervention group than in the control group would lead to uncertainty as to whether any observed effect resulted from the difference in the sodium intake between the two groups. Providing a “lead-in”²¹ diet or a “wash-out” period²² for studies with a cross-over design for an adequate duration prior to randomization can help reduce the likelihood of different baseline values.

- *How were the results from the intervention and control groups statistically analyzed?* Statistical analysis of the study data is a critical factor because it provides the comparison between subjects consuming the substance and those not consuming the substance, to determine whether there is a reduction in risk of the disease. Furthermore, when conducting statistical analyses among more than two groups, the data should be analyzed by a test designed for multiple comparisons (e.g., Bonferroni, Duncan). Thus when statistical analyses are not performed between the control and intervention group or are conducted inappropriately, scientific conclusions cannot be drawn about the role of the substance in reducing the risk of the disease and, therefore, the agency does not intend to use such studies to evaluate the substance/disease relationship.

- *What type of biomarker of disease risk was measured?* As discussed above, when the study does not measure disease incidence or associated mortality, then surrogate endpoints are essential for measuring risk. Scientific conclusions cannot be drawn about the relationship between the substance and risk of the disease if the risk biomarker is not a surrogate endpoint (see discussion above in Section III.C). The agency does not intend

²¹ A diet that is provided to all study groups prior to randomization.

²² Time period within a cross-over design study during which subjects do not receive an intervention.

Contains Nonbinding Recommendations

to use such studies from which scientific conclusions cannot be drawn in its evaluation of the substance/disease relationship.

- *How long was the study conducted?* Studies that use a surrogate endpoint should be conducted long enough to ensure that any change in the endpoint is in response to the dietary intervention. If the study is run for a short time period such that the effects of the substance cannot be evaluated, then scientific conclusions cannot be drawn about the relationship between the substance and the disease and, therefore, the agency does not intend to use such a study to evaluate the substance/disease relationship. For example, FDA has considered 3 weeks to be the minimum duration for evaluating the effect of an intervention with various saturated fats on serum LDL cholesterol concentration (Kris-Etherton and Dietschy, 1997)

- *If the intervention involved dietary advice, was there proper follow-up to ascertain whether the advice resulted in altered intake of the substance?* When the dietary intervention involves dietary advice rather than a prescribed diet administered under a controlled condition, there should be some type of assessment of the changes in intake of the substance (e.g., dietary assessment or measurement of a biomarker of intake in response to dietary advice). Without some type of assessment of whether the dietary advice resulted in a change in intake of the substance, scientific conclusions cannot be drawn about the substance/disease relationship and, therefore, the agency does not intend to use studies that lack such an assessment to evaluate the substance/disease relationship.

- *Where were the studies conducted?* It is important that the study population is relevant to the general U.S. population or the population subgroup identified in the proposed claim. Thus, FDA evaluates each study to determine if the study population lives in an area where malnutrition or inadequate intakes of the specific substance is common, and/or where the prevalence or etiology of the disease that is the subject of the claim is not similar to that in the United States. For certain countries, there may be risk factors of a specific disease that are not relevant to disease risk in the United States (e.g., risk factors for gastric cancer in certain Asian countries). Differences in nutrition, diet, and disease risk factors between the United States and the country where a study was done may mean that the study results cannot be extrapolated to the U.S population or population subgroup. For example, scientific conclusions about the comparatively well-nourished U.S. population cannot be drawn from studies in subjects that are malnourished. Nutrient status and metabolism can be severely altered when an individual is malnourished, and therefore the effect of the substance on a particular surrogate endpoint may be very different between a malnourished and well-nourished individual (Shils et al., 2006). Scientific conclusions cannot be drawn from studies conducted in countries or regions where inadequate intake of the substance is common since a response to the intake of the substance may be due to the correction of a nutrient deficiency for which health claims are not intended.

Furthermore, conclusions cannot be drawn from studies conducted in countries or regions where the etiology of the disease is very different than in the United States. For example, major risk factors for gastric cancer in Japan (high salt intake and *Helicobacter pylori* (H.

Contains Nonbinding Recommendations

pylori) infection) are significantly more prevalent than in the United States. Therefore, it is not appropriate to extrapolate from data on a Japanese population concerning the relationship between a substance and gastric cancer to reach conclusions about potential effects on the U.S. population.

Observational Studies

- *What type of information was collected?* Biological samples (e.g., blood, urine, tissue, or hair) should be used to establish intake of a substance only if a dose-response relationship has been demonstrated between intake of the substance and the level of the substance (or a metabolite of the substance) in the biological sample. There should be evidence to demonstrate a strong correlation²³ between the intake level of the substance and the level of the substance or a metabolite in the biological sample (e.g., selenium intake and serum selenium concentration). If the correlation is weak for a specific biological sample, then scientific conclusions cannot be drawn from studies that used that biological sample as a biomarker of intake. Biological samples in case-control studies should not be used to establish intake of the substance since the metabolism or concentration of the substance may be altered in subjects as a result of the disease.
- *Were scientifically acceptable and validated dietary assessment methods used to estimate intake of the substance?* A single 24-hour diet recall or diet record is generally regarded as an inadequate method for assessing an individual's usual intake of a substance, although it may be useful for assessing mean intake of a group. A diet history involves extensive interviews with the study subjects. However, diet histories are also usually inadequate for assessing intake of a substance since respondents are asked to make judgments about intakes of usual foods and the amounts eaten. A food frequency questionnaire contains a limited number of food items and is inadequate for assessing intake of a substance if the major sources of the substance are not included in the questionnaire. Food frequency questionnaires also do not always account for different varieties of a particular food or different cooking methods. Because of these limitations, validation of the food frequency questionnaire method to assess food intake is essential in order to be able to draw conclusions from the scientific data, as the failure to validate may lead to false associations between dietary factors and diseases or disease-related markers.²⁴

²³ Correlation is evaluated using correlation coefficients (r). Correlation coefficients range from -1 (negative correlation) through +1 (positive correlation). The closer to 1, the stronger the correlation; the closer to zero, the weaker the correlation.

²⁴ "Validation of the food frequency questionnaire method is essential, as incorrect information may lead to false associations between dietary factors and disease or disease-related markers." Cade, J., Thompson, R., Burley, V., and Warm D. Development, Validation and Utilization of Food-Frequency Questionnaires-A Review. *Public Health Nutrition*, 5: page 573, 2002. See, also, Subar, A., et al., Comparative validation of the Block, Willett, and National Cancer Institute Food Frequency Questionnaires, *American Journal of Epidemiology*, 154: 1089-1099, 2001.

Contains Nonbinding Recommendations

• *Did the observational study evaluate the relationship between a disease and a food or a food component?* Because observational studies estimate intake of a whole food based on recorded dietary intake methods such as food frequency questionnaires, diet recalls, or diet records, a common weakness of observational studies is the limited ability to ascertain the actual intake of the substance for the population studied. Furthermore, if the substance is a food component rather than a whole food, there is an additional estimation of the amount of the food component that is present in the individual foods. The content of foods' components can vary based on factors such as soil composition, food processing/cooking procedures, or storage (duration, temperature). Thus, it is difficult to ascertain an accurate amount of the food component consumed based on reports of dietary intake of whole foods.

In addition, the whole food and products that include several food components, e.g., multi-nutrient dietary supplements, contain not only the food component that is the subject of the claim, but also other food components that may be associated with the metabolism of the food component of interest or the pathogenesis of the disease or health-related condition. Because whole foods and products such as multi-nutrient dietary supplements consist of many food components, it is difficult to study the food components in isolation (Sempos et al., 1999). For studies based on recorded dietary intake of whole foods or multiple food components, it is not possible to accurately determine whether any observed effects of the food component that is the subject of the claim on disease risk were due to: (1) that food component alone; (2) interactions with other food components; (3) other food components acting alone or together; or (4) decreased consumption of other substances contained in foods displaced from the diet by the increased intake of foods rich in the food component of interest (See Sempos et al. (1999), Willett (1990) and Willett (1998) regarding the complexity of identifying the relationship between a specific food component within a food and a disease).

In fact, evidence demonstrates that in a number of instances, observational studies based on the recorded dietary intake of conventional foods may indicate a benefit for a particular nutrient with respect to a disease, but it is subsequently demonstrated in an intervention study that the nutrient-containing dietary supplement does not confer a benefit or actually *increases* risk of the disease (Lichtenstein and Russell, 2005). For example, previous observational studies reported an association between fruits and vegetables high in beta-carotene and a reduced risk of lung cancer (Peto et al., 1981). However, subsequent intervention studies, the Alpha-Tocopherol and Beta Carotene Prevention Study (ATBC) and the Carotene and Retinol Efficiency Trial (CARET), demonstrated that beta-carotene supplements increase the risk of lung cancer in smokers and asbestos-exposed workers, respectively (The Alpha-Tocopherol and Beta Carotene Cancer Prevention Study Group, 1994; Omenn et al., 1996). These studies illustrate that the effect of a nutrient provided as a dietary supplement exhibits different health effects compared to when it is consumed among many other food components. Furthermore, these studies demonstrate the potential public health risk of relying on results from epidemiological studies, in which the effect of a nutrient is based on recorded dietary intake of conventional foods as the sole source for concluding that a relationship exists between a specific nutrient and disease risk; the effect could actually be harmful. For the

Contains Nonbinding Recommendations

above reasons, scientific conclusions from observational studies cannot be drawn about a relationship between a food component and a disease. Observational studies, however, can be used to measure associations between a whole food and a disease.²⁵

E. Assessing the Methodological Quality of Studies

For the studies that are not eliminated during the earlier evaluation, FDA intends to independently rate each such study for methodological quality. Studies can receive a high, moderate, or low quality rating. FDA intends to base this quality rating on several factors related to study design, data collection, the quality of the statistical analysis, the type of outcome measured, and study population characteristics other than relevance to the U.S. population (e.g., selection bias and the provision of important subject information [e.g., age, smokers]). If the scientific study adequately addressed all or most of the above factors, FDA plans to give it a high methodological quality rating. FDA plans to give moderate or low quality ratings based on the extent of the deficiencies or uncertainties in the quality factors. Studies that are so deficient in quality that they receive a low quality rating are studies from which scientific conclusions cannot be drawn about the substance/disease relationship and are eliminated from further review.

Examples of factors FDA intends to consider in assessing the methodological quality of individual studies remaining at this point in the scientific evaluation approach set out in this guidance include the following:

²⁵ In *Pearson v. Shalala*, the D.C. Circuit noted that FDA had "logically determined" that the consumption of antioxidant vitamins in dietary supplement form could not be scientifically proven to reduce the risk of cancer where the existing research had examined only foods containing antioxidant vitamins, as the effect of those foods on reducing the risk of cancer may have resulted from other substances in those foods. 164 F.3d 650, 568 (D.C. Cir. 1999). The D.C. Circuit, however, concluded that FDA's concern with granting antioxidant vitamins a qualified health claim could be accommodated by simply adding a prominent disclaimer noting that the evidence for such a claim was inconclusive, given that the studies supporting the claim were based on foods containing other substances that might actually be responsible for reducing the risk of cancer. *Id.* The court noted that FDA did not assert that the dietary supplements at issue would "threaten consumer's health and safety." *Id.* at 656. There is, however, a more fundamental problem with allowing qualified health claims for individual nutrients based on studies of foods containing those nutrients than the problem the D.C. Circuit held could be cured with a disclaimer. Even if the effect of the specific component of the food could be determined with certainty, recent scientific findings on the complex nature of nutrient-food interactions and on the relationships among diet, biological parameters, and disease indicate that nutrients found to have health benefits when consumed in one food or group of foods may not necessarily have the same beneficial effect when they are consumed in dietary supplement form or in other foods. See Lichtenstein and Russell (2005). For example, not only have studies on dietary supplements established that the benefits associated with the dietary intake of certain nutrients do not materialize when the nutrients are taken as a supplement, but some of these studies have actually indicated an increased risk for the very disease the nutrients were predicted to prevent. *Id.* Thus, a study based on intake of a specific food or foods provides no information from which scientific conclusions may be drawn for the nutrient itself. Further, even if the nutrients are consumed in other foods rather than in a dietary supplement, the physiological effects may be different because the food matrix can affect the bioavailability and bioactivity of the nutrients. *Id.*

Contains Nonbinding Recommendations

Intervention Studies

- *Were the studies randomized and blinded and was a placebo provided?* Appropriate randomization eliminates intrinsic and/or extrinsic factors, other than the substance, that could have an influence on the outcome of the study. Blinding is especially important when the endpoint can be influenced by a subject's awareness that he or she is receiving something that may be beneficial. Blinding would be critical when the outcome measure is cognitive performance, mental status (e.g. memory, depression), or behavior. Including a placebo in a supplementation trial prevents the subject from knowing whether he or she is receiving the substance or not.
- *Were inclusion/exclusion criteria and key information on the characteristics of the study population provided?* For instance, were healthy or high-risk subjects allowed to take medications that can affect the disease that is subject of the claim during the study? If so, was the proportion of subjects taking medications similar between the control and intervention groups?
- *Was subject attrition (subjects leaving the study before the study is completed) assessed, explained in the article reporting the study, and reasonable?* If there were a marked number of drop-outs, then it would be important to know why subjects dropped out and how the drop-outs affected the number and composition of the intervention and placebo group.
- *How was compliance with the study protocol verified?* Intervention studies should include a mechanism for verifying that the subjects followed the study protocol. For example, a supplementation trial should have a mechanism for determining how frequently the subjects took their supplements. It would be important to know 1) if the subjects took all of the supplements provided by the study or only a portion and 2) what proportion of the subjects for each group took less than the directed amount.
- *Was statistical analysis conducted on baseline data for the all subjects initially enrolled in the study or only those who completed the study?* If there were a marked number of drop-outs which, in turn, affected the composition of the intervention groups differently from the placebo groups, then it would be important to determine if statistical analysis on baseline data was conducted for all subjects initially enrolled in the study or only for those who completed the study.
- *Did the study measure disease incidence or a surrogate endpoint of disease risk?* While surrogate endpoints of disease risk have been validated, they are not as accurate as measuring the actual onset of a disease. This quality issue would also apply to observational studies.
- *How was the onset of a disease determined?* When disease incidence is the endpoint being measured, it is important that the disease that is subject of the claim is confirmed either through medical records and/or pathology reports. Relying on less specific records,

Contains Nonbinding Recommendations

such as death certificates, is not sufficient. This quality issue would also apply to observational studies.

Observational Studies

- *Was there an adequate adjustment for confounders of disease risk?* Several aspects of a substance/disease relationship may give rise to confounders. Therefore, it is important to adjust for confounders of the disease of interest so that observed effects on risk of disease that may be due to confounders are not incorrectly attributed to the substance of interest. For example, there can be multiple non-dietary risk factors for a disease (e.g., smoking, body mass index, and age for hypertension). Therefore, when evaluating the relationship between sodium and blood pressure, an adjustment of the risk analysis should be made based on age, smoking, body mass index and age.
- *What type of dietary assessment method was used to estimate dietary intake?* Validated food frequency questionnaires are more reliable in estimating “usual” intake of foods compared to diet records or 24-hour recall methods. See Section III.B.

F. Evaluating the Totality of Scientific Evidence

Under the approach set out in this guidance, at this point, FDA intends to evaluate the results of the studies from which scientific conclusions can be drawn and rate the strength of the total body of publicly available evidence. The agency plans to conduct this evaluation by considering the study type (e.g., intervention, prospective cohort, case-control, cross-sectional), methodological quality rating previously assigned, number of the various types of studies and sample sizes, relevance of the body of scientific evidence to the U.S. population or target subgroup, whether study results supporting the proposed claim have been replicated²⁶, and the overall consistency^{27,28} of the total body of evidence. Based on the totality of the scientific evidence, FDA determines whether such evidence meets the SSA standard or whether such evidence is credible to support a qualified health claim for the substance/disease relationship.

²⁶ Replication of scientific findings is important for evaluating the strength of scientific evidence (Wilson, E.B. *An Introduction to Scientific Research*. Dover Publications, 1990; pages 46-48).

²⁷ In this guidance, “consistency” is used to mean the level of agreement among the studies from which scientific conclusions could be drawn about the substance/ disease relationship.

²⁸ Consistency of findings among similar and different study designs is important for evaluating causation and the strength of scientific evidence (Hill A.B. *The environment and disease: association or causation? Proceedings of the Royal Society of Medicine*. 1965;58:295-300).; see also Agency for Healthcare Research and Quality, *Systems to Rate the Scientific Evidence*, which defines “consistency” as “the extent to which similar findings are reported using similar and different study designs.”
[<http://www.ahrq.gov/clinic/epcsums/strengthsum.htm#Contents>]

Contains Nonbinding Recommendations

Within each study type, the studies are reviewed for:

- *Number* of studies and number of subjects per group
- *Methodological quality* (high, moderate, or low).
- *Outcome* (beneficial effect, no effect, adverse effect) of the studies within each study type. For the outcome of an intervention study to demonstrate an effect, the intervention group should be statistically significantly different from the control group ($P < 0.05$). For observational studies, confidence intervals (CI) for risk are significant when the value is less than or greater than “1”. Many studies analyze for the statistical significance of the linear relationship (P for trend) between the substance and the disease. While this trend may be significant ($P < 0.05$), the difference in risk between subjects at the various levels of intake (e.g., tertiles, quartiles or quintiles of intake)²⁹ may not be significant. In that case, the studies show no effect. Evaluation of the size of the effect (e.g., percent reduction in LDL cholesterol) may be useful for comparing effects within a study (e.g., relative effect of two forms of the substance or the relative effect of frequency of consumption).
- In general, the greater the *consistency* among the studies in showing a beneficial relationship, the greater the level of confidence that a substance/disease relationship exists. Conflicting results do not disprove an association (because the elements of the study design may account for the lack of an effect in negative studies) but tend to weaken confidence in the strength of the association. The greater the magnitude of the beneficial effect, the more likely the association may exist.
- *Relevance* to the general U.S. population. For example,

To what extent did the studies that showed a benefit include populations that represent the general U.S. population or a population subgroup (e.g., elderly, women)?

Did the studies only include subjects with unique lifestyles (e.g., smokers, vegetarians)?

Do the studies suggest that the intake level of the substance that provides a benefit significantly exceeds usual intakes in the United States?

²⁹ Tertile, quartile and quintile of intake is the result of dividing a study population into 3, 4 or 5 groups, respectively, such that the average intake level of the substance varies across the groups (e.g., lowest intake group represents the lowest tertile of intake and the highest intake group represents the highest tertile). The study population is divided such that each group has the same number of subjects.

Contains Nonbinding Recommendations

FDA evaluates whether the totality of the evidence supports a claim for the entire U.S. population or just a subgroup. If the evidence only supports a claim for a subgroup, that information would be set out in the claim. If the substance is one that must be used for risk reduction at much higher levels than the normal U.S. intake, that information would also be reflected in the claim.

In general, intervention studies provide the strongest evidence for the claimed effect, regardless of existing observational studies on the same relationship. Intervention studies are designed to avoid selection bias and avoid findings that are due to chance or other confounders of disease (Sempos et al., 1999). Although the evaluation of substance/disease relationships often involves both intervention and observational studies, observational studies generally cannot be used to rule out the findings from more reliable intervention studies (Sempos et al., 1999). One intervention study would not be sufficient to rule out consistent findings of observational studies. However, when several randomized, controlled intervention studies are consistent in showing or not showing a substance/disease relationship, they trump the findings of any number of observational studies (Barton, 2005). This is because intervention studies are designed and controlled to test whether there is evidence of a cause and effect relationship between the substance and the reduced risk of a disease, whereas observational studies are only able to identify possible associations. There are numerous examples -- such as vitamin E and CVD and beta-carotene and lung cancer -- where associations identified in observational studies have been publicized. However, when randomized, controlled intervention studies were later conducted to test these possible associations, the intervention studies found no evidence to support the relationships (Lichtenstein and Russell, 2005).

G. Assessing Significant Scientific Agreement

Significant scientific agreement refers to the extent of agreement among qualified experts in the field. Health claims represent a continuum of scientific evidence that extends from very limited or inconclusive evidence to consensus, with evidence supporting SSA health claims lying closer to consensus. FDA's determination of SSA represents the agency's best judgment as to whether qualified experts would likely agree that the scientific evidence supports the substance/disease relationship that is the subject of a proposed health claim. The SSA standard is intended to be a strong standard that provides a high level of confidence in the validity of the substance/disease relationship. SSA means that the validity of the relationship is not likely to be reversed by new and evolving science, although the exact nature of the relationship may need to be refined. SSA does not require a consensus based on unanimous and incontrovertible scientific opinion. SSA occurs well after the stage of emerging science, where data and information permit an inference, but before the point of unanimous agreement within the relevant scientific community that the inference is valid.

For qualified experts to reach an informed opinion regarding the validity of a claim, the data and information that pertain to the claim must be available to the relevant scientific community. A finding of SSA then derives from the conclusion that there is a sufficient

Contains Nonbinding Recommendations

body of relevant, publicly available scientific evidence that shows consistency across different studies and among different researchers. The usual mechanism to show that the evidence is available to qualified experts is that the data and information are published in peer-reviewed scientific journals. The value of an expert's opinion will be limited if he or she did not have access to all the evidence.

In determining whether there is significant scientific agreement, FDA takes into account the viewpoints of qualified experts outside the agency, if evaluations by such experts have been conducted and are publicly available. For example, FDA intends to take into account:

- documentation of the opinion of an "expert panel" that is specifically convened for this purpose by a credible, independent body;
- the opinion or recommendation of a federal government scientific body such as the National Institutes of Health (NIH) or the Centers for Disease Control and Prevention (CDC); or the National Academy of Sciences (NAS);
- the opinion of an independent, expert body such as the Committee on Nutrition of the American Academy of Pediatrics (AAP), the American Heart Association (AHA), American Cancer Society (ACS), or task forces or other groups assembled by the National Institutes of Health (NIH);
- review publications that critically summarize data and information in the secondary scientific literature.

FDA accords the greatest weight to the conclusions of federal government scientific bodies, especially when the evidence for the validity of a substance/disease relationship has been judged by such a body to be sufficient to justify dietary recommendations to the public. When the validity of a substance/disease relationship is supported by the conclusions of federal government scientific bodies, FDA typically finds that significant scientific agreement exists. Conclusions of other expert bodies may also be relevant to support a determination of SSA. Although reviews by individual outside experts are considered in assessing SSA, evidence from such reviews alone would not necessarily support a conclusion that the standard has been met, especially if the conclusions of such reviews were not supported by available assessments of the same body of evidence from federal scientific bodies, expert panels, or independent expert bodies. Reviews by outside experts or expert panels are most useful when there is a reasonable basis to conclude that they represent the larger group of qualified experts in the field. Most importantly, the relevance of an outside expert review depends on whether the evidence examined applies to the claim in terms of considerations such as specification and measurement of the substance and the disease.

When conclusions from qualified experts are not available (for instance, if the data supporting a proposed health claim are relatively new and have not yet been reviewed by an independent expert panel or body), a compelling and relevant body of evidence may

Contains Nonbinding Recommendations

nonetheless cause the agency to conclude that significant scientific agreement exists. Because each situation may differ with the nature of the claimed substance/disease relationship, it is necessary to consider both the extent of agreement and the nature of the disagreement on a case-by-case basis. If scientific agreement were to be assessed under arbitrary quantitative or rigidly defined criteria, the resulting inflexibility could cause some valid claims to be disallowed where the disagreement, while present, is not persuasive.

Application of the significant scientific agreement standard is intended to be objective, in relying upon a body of sound and relevant scientific data; flexible, in recognizing the variability in the amount and type of data needed to support the validity of different substance/disease relationships; and responsive, in recognizing the need to re-evaluate data over time as research questions and experimental approaches are refined.

H. Specificity of the Claim Language for Qualified Health Claims

When the evidence for a substance-disease relationship is credible but does not meet the SSA standard, then the proposed claim for the relationship should include qualifying language that expresses the level of scientific evidence to support the relationship.

The health claim language should reflect the level of scientific evidence with specificity and accuracy. However, gaps in the scientific evidence may sometimes limit the information that can be included in the claims. For example, when the scientific evidence is limited but credible, it may not be possible for the qualified health claim to identify an amount of the substance that is associated with a reduced risk of the disease.

Under FDA's health claim regulations, a health claim must specify the daily dietary intake of the substance necessary to achieve the claimed effect when there is no regulation defining what constitutes a "high" level of the substance in food (21 CFR 101.14(d)(2)(vii)). FDA has defined "high" in its nutrient content claim regulations as meaning that the food contains 20% or more of the Daily Value of the substance (21 CFR 101.54(b)). Therefore, when no Daily Value for the substance has been established, the agency cannot establish a definition for a "high" level of the substance. When the substance that is the subject of the claim has no Daily Value, FDA determines the daily dietary intake necessary to achieve the claimed effect whenever the available evidence is sufficient to make such a determination possible. See, e.g., 21 CFR 101.83(c)(2)(G) (health claim regulation for plant sterol/stanol esters and reduced risk of coronary heart disease). However, there are times when the credible evidence for the risk reduction effect is not specific enough for FDA to identify even a possible level of intake for the general U.S. population. See FDA's September 8, 2004, letter of enforcement discretion for qualified health claim about omega-3 fatty acids and reduced risk of coronary heart disease (Martek petition) (<http://www.cfsan.fda.gov/~dms/ds-ltr37.html>).

Contains Nonbinding Recommendations

When there is credible evidence available to suggest a relationship between the substance and disease, it is important to determine whether the substance has an independent role in the relationship or whether its role is based on the inclusion or replacement (i.e., substitution) of other substances. An example of where the evaluation of the independent role of a substance can be challenging is when the substance is a conventional food or macronutrient (e.g., fat or carbohydrate). In studies evaluating the possible health effects of a conventional food or macronutrient, the inclusion of either in the diet usually requires the removal of other conventional foods or macronutrients (i.e., substitution to yield isocaloric diets). If it is determined that the substance does not play an independent role and/or requires the reduction or inclusion of another substance to show a beneficial effect, the claim language will reflect this finding

I. Reevaluation of Existing SSA or Qualified Health Claims

FDA may reevaluate a health claim in response to a petitioner or on its own initiative, and when it does so it intends to use the scientific evaluation process described above. To maximize the public health benefit of its health claims review, FDA intends to evaluate new information that becomes available to determine whether it necessitates a change to an existing SSA or qualified health claim. For example, scientific evidence may become available that will (1) support the revision of claim language for an SSA or qualified health claim, (2) support change of an SSA claim to a QHC or support change of a QHC to an SSA claim, or (3) raise safety concerns about the substance that is the subject of a health claim or otherwise no longer support a health claim (SSA or QHC).

IV. References

American Cancer Society, Cancer Facts and Figures, 2004.

The Alpha-Tocopherol, Beta Carotene Cancer Prevention Study Group. The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. *New England Journal of Medicine* 1994; 330:1029-1035.

Barton S. Which clinical studies provide the best evidence? The best RCT still trumps the best observational study. *British Medical Journal* 2000; 321:255-256.

Cade J, Thompson R, Burley V, Warm D. Development, validation and utilization of food-frequency questionnaires – a review. *Public Health Nutrition* 2002; 5:567-587.

Federal Judicial Center, Reference Manual on Scientific Evidence, Second Edition, 2000.

Flegal KM. Evaluating epidemiological evidence of the effects of food and nutrient exposures. *American Journal of Clinical Nutrition* 1999; 69:1339S-1344S.

Contains Nonbinding Recommendations

Greer N, Mosser G, Logan G, Halaas GW. A practical approach to evidence grading. *Joint Commission Journal on Quality Improvements* 2000; 26:700-712.

Hill AB. The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine* 1965; 58:295-300

Hord NG, Fenton JJ. Context is everything: mining the normal and preneoplastic microenvironment for insights into the diet and cancer risk conundrum. *Molecular Nutrition and Food Research* 2007; 51:100-106.

IOM, Institute of Medicine. Dietary Supplements: A Framework for Evaluating Safety. National Academies Press, Washington, DC. 2005.

Kraemer HC, Lowe KK, Kupfer DJ. To Your Health: How to Understand What Research Tell Us About Risk. Oxford University Press, 2005.

Kris-Etherton PM, Dietschy J. Design criteria for studies examining individual fatty acid effects on cardiovascular disease risk factors: human and animal studies. *American Journal of Clinical Nutrition* 1997; 65:1590S-1596S.

Lichtenstein AH, Russell RM. Essential Nutrients: Food or Supplements? *Journal of American Medical Association* 2005; 294:351-358.

Milner JA. Diet and Cancer: Facts and Controversies. *Nutrition and Cancer* 2006; 56: 216-224.

National Cancer Institute, Dictionary of Cancer Terms,
<http://www.cancer.gov/dictionary>

National Research Council. Diet and Health: Implications for Reducing Chronic Disease Risk. National Academy Press, Washington, DC, 1989.

Omenn, GS, Goodman GE, Thornquist MD, Balmes J, Cullen MR, Glass A, Keogh JP, Meyskens FL, Valanis B, Williams JH, Barnhart S, Hammer S. Effects of a combination of beta carotene and vitamin A on lung cancer and cardiovascular disease. *New England Journal of Medicine* 1996; 334:1150-1155.

Peto R, Doll R, Buckley JD, Sporn MB. Can dietary beta-carotene materially reduce human cancer rates? *Nature* 1981; 290:201-208.

Sempos CT, Liu K, Earnst ND. Food and nutrient exposures: what to consider when evaluating epidemiologic data. *American Journal of Clinical Nutrition* 1999; 69:1330S-1338S.

Torun B. Protein-energy malnutrition. In: Modern Nutrition in Health and Disease. Williams and Williams, New York, 2006.

Contains Nonbinding Recommendations

Spilker B. Guide to Clinical Studies. Raven Press, New York, 1991.

Subar AF, Thompson FE, Kipnis V, Midthune D, Hurwitz P, McNutt S, McIntosh A, Rosenfeld S. Comparative validation of the Block, Willett, and National Cancer Institute Food Frequency Questionnaires, *American Journal of Epidemiology* 2001; 154: 1089-1099.

Szklo M., Nieto FJ. Epidemiology Beyond the Basics, Aspen Publishing, 2000.

Willett W.C. Overview of nutritional epidemiology. Nutritional Epidemiology, Oxford University Press, Oxford. 1990.

Willett W.C. Issues in analysis and presentation of dietary data. In: Nutritional Epidemiology, Second Edition, Oxford University Press, Oxford, 1998.

Wilson E.B. An Introduction to Scientific Research, General Publishing Company, Toronto, 1990.